DIGITAL INVESTIGATION LEGAL EDITORIAL

The Internet is the computer: The role of forensics in bridging the digital and physical divide

Just as residue from the ridge patterns on our fingers existed before science and technology was able to "uncover" them by latent fingerprinting methods, digital traces of criminal activities exist on the Internet, and consequently lie dormant because we are only beginning to develop and deploy tools and techniques to identify, manage, and model solutions to probing questions related to digital crimes.

Cyber forensic investigations\(^1\) occur in varying degrees throughout the fields of computer security and incident response, network forensics, and law enforcement investigations, yet in all contexts involve the recognition, recovery and reconstruction of investigatory leads and evidence. In the context of investigations, the source-artifacts for evidence and leads are often "silied" into investigation data from structured law enforcement reports, or data from examinations of computers involved in a crime. No longer are artifacts relegated to data in reports, and no longer is the stand-alone computer exclusively a target or tool used in criminal activity. The Internet itself has become a breeding ground for primary and secondary sources of evidence in the search for truth, as well as harboring the seeds to predict future malfeasance.

Like other forensic sciences, fundamental methods of cyber forensics begin by collecting a large number of intensely diverse variables or attributes, and culminate in pattern matching among these variables to individualize evidence.

Computer security and incident response, and network forensics increasingly require digital investigations involving linking heterogeneous data sets that contain remnants of human activity, often times occurring across multiple environments. Evidence identification and pattern matching in this context consists of recognizing and correlating digital artifacts contained within and among various data sources such as Web pages, computer logs, Internet newsgroups, online chat rooms, and corporeal case reports — each with different levels of granularity and context. Nevertheless, linkage of this data is becoming more important for the efficient administration of justice in a 21st Century society that is increasingly conducting its collective life in the digital realm.\(^2\)

What is our challenge?

Law enforcement (LE) is an information-intensive process in which government agencies are called upon to collect and interpret large public data sets in an effort to serve and protect the citizenry, while at the same time maintain trust and

---

\(^1\) For the purposes of this article, "cyber forensics" is used by the author to refer to the novel subcategory of "Internet forensics", defined as repeatable techniques and methodologies to collect, preserve and analyze digital data on the Internet for investigation purposes. Note that, "computer forensics" is the principle applied to the collection, preservation and analysis of computer-derived evidence to ensure its admissibility as evidence in a legal proceeding.

\(^2\) E-commerce, email and VOIP (voice-over-Internet-protocol) communications are a few prominent examples of the ubiquity of computer-based transactions in modern society.
reliability in fulfilling its mission. However, LE is by its very nature reactionary to information contained within and derived from reports of criminal activity. As a result, the effectiveness of LE is directly related to the quality of information reported and proficiency of the subsequent analyses. The process of law enforcement has thus far encountered technical, managerial and socio-legal barriers to integrating, correlating and interpreting intra-agency crime data with public, Internet-based data. The challenge lies in developing tools and methodologies to enhance the forensic soundness of the evidentiary and investigation artifacts.\(^3\)

One of the most prevalent challenges facing LE in our information society is to integrate public, Internet-based data with existing private data sets to enhance its duty to enforce laws as well as its mission to protect and serve the public citizenry. Fulfilling this expectation in isolation from other law enforcement entities and public data sources is no longer tenable, especially in light of information technology advances and pressure to enhance predictive capabilities. Although there have been a handful of approaches that allow inter-agency and inter-jurisdiction law enforcement data, constructing new approaches that expand this data correlation to encompass public, Internet-based data to produce forensically sound and actionable information is a mounting priority. The onslaught of identity theft, phishing and other Internet fraud crimes are indeed driving the need to bridge the physical and digital gap.

The information-driven nature of law enforcement begins with initial data collection at the crime scene or via victim reporting, extends through evidence and intelligence gathering, and culminates in analyses of data to support the prosecution and aid in preventing criminal activities. However, LE is by its very nature reactionary to information contained within and derived from reports of criminal activity. As a result, the effectiveness of law enforcement is directly related to the quality of information reported and proficiency of the subsequent analyses. Collecting, processing, organizing and analyzing reports between agencies enhance this quality.

Nevertheless, there is a chasm between information contained in crime-related reports and the forensically-relevant (who, what, when, where, how, why) data that exists independent of crime reports. The breadth of forensically-relevant data available on the Internet can impact the quality of actionable information contained within existing private records maintained by LE.

### Technical, managerial and socio-legal factors

The Internet has emerged as a mainstream vehicle for global communications among persons, informal groups and public organizations, corporations and governments. Billions of pages of information have been posted to the Internet using various protocols, including http, irc/irc (Internet relay chat/direct client communications), ftp (file transfer protocol), Usenet (newsgroups), auctions and peer-to-peer services. Unfortunately, the same characteristics that have made the Internet so attractive for business and government — low cost, high-speed, low barrier to entry, multi-media capable, etc. — have also made it highly compelling for fraudsters, terrorists and organized criminal groups.

LE's exploitation of the Internet as an intelligence and investigative resource has been complicated by the lack of a readily trained cadre of government collectors and analysts, established operational processes and an accessible collection of analysis platforms capable of supporting high-volume content collection, reduction, aggregation, analysis, reporting and assessment. The lack of such a platform has relegated cyber forensics to ad hoc, "hit-or-miss" efforts. Methodical, systematic collection has typically not been performed.

We need to employ technologies to increase the scope and quality of the information upon which LE depends to ensure the public safety. As the persons engaging in unlawful activities grow increasingly dependent on the Internet as a tool (i.e. communication or transaction mechanism) and/or target (i.e. the use of the Internet to commit new crimes or old crimes in new ways) to facilitate their offences they leave a trail of evidence and investigatory artifacts as a natural by-product. From this vast and disparate well of publicly accessible data much can be uncovered and inferred.

---

\(^3\) To address this need, the author is engaged in several research projects related to developing the methodologies and infrastructure for the management, analysis and presentation of public and private multidimensional data. This includes addressing legal, policy, technical and analytical models, methods, tools and techniques to effectively integrate and correlate law enforcement information with public, cyber-based information to study the impact of this expanded dimensional information on the efficient remediation and proactive capabilities of law enforcement.
Currently, the search, collection, and analysis of information evidence from the public Internet have been relegated to a few, highly specialized, and usually grant-funded law enforcement projects. It has been addressed, conceptually, as a new kind of high technology criminal problem. It is indeed that, but also much more. Cyberspace has become the neighborhood wherein law enforcement officers must regularly interact with their constituency. The previously specialized projects and investigation techniques confined to the non-budgeted criminal justice arena will certainly be deployed universally. Regular law enforcement reporting and records systems, evidence collection and analysis systems, and proactive crime suppression activities must likewise be relocated to the virtual existence of cyberspace and scaled well beyond traditional jurisdictional barriers. The alternative is failure of law in society.

Integrating and/or correlating public data with justice data will assist in gathering and exchanging information that can provide the empirical data needed to assess various threats. The frequency (number of occurrences) and severity data may not find their way into LE reports because formal cyber reporting requirements are nascent, but nonetheless, information needed to provide a meaningful assessment may certainly lurk informally on the Net. Furthermore, cases involving events related to the investigation at hand may be inaccessible or unverifiable for reasons ranging from out of court settlements to unpublished opinions.

Indeed, traditional investigations that index on the individual, rather than querying on the criminal event itself may be problematic. This is where large-scale forensic analytics and modeling can significantly enhance the ability to infer behaviour and intent from patterns of activity (i.e. cyber crime usage signatures). For instance, a query on a particular suspect who may perpetrate identity theft using the Internet under multiple aliases will largely fly under the radar of traditional investigations that do not make use of Internet-based information.

While models of certain types of criminals have been available for sometime, models of how these criminals may utilize the Internet to commit transgressions, as well as composites of persons committing cyber-based crimes are very immature (Computer Science and Telecommunications Board, 2000). Efforts should be directed at documenting, collecting, and memorializing this data. A corollary challenge lies in compiling the data and constructing the models needed for correlating activity whether the approach is a centralized repository or distributed federation of data.

While the Internet offers the benefit of a new dimension of information and unprecedented ability to interact with remote groups, it is not without potential dangers that must be considered in any investigation management, analysis and processing model. For example, the reliability and credibility of the linkage between Internet-based data and current corporal-based report events may need some degree of quantification and qualification. To explain further, the collection and entry of crime report data are facilitated by trained officers who observe traditional interactions where issues of competence, coercion, malice, or willingness are effectively evaluated. However, these traditional metrics may not be present when considering information obtained from impersonal computer-to-computer interaction on the Internet. Furthermore, there are important privacy ramifications of intermingling data with various reliability and quality standards, exacerbated by a relative dearth of clear policies or guidelines. In short, measuring the context of cyber data is a growing challenge.

Conclusion

Cyber crime will not 'cease and desist' in deference to LE’s ability to utilize the artifacts it leaves behind. To be sure, the rapid pace of technology development has not only fostered new criminal acts (i.e. the spread of computer virii, unauthorized access to a computer system, possession of access control devices, etc.), but in fact has spawned novel means to conduct traditional crimes (i.e. online fraud, Internet gambling, copyright infringement, etc.).

Therefore, the question is not whether evidence exists, but rather, whether LE can uncover, contextualize and integrate cyber evidence with predication data from traditional case reports. Our focus as practitioners in this arena is to support the advancement of tools and techniques to manage, analyze and present public and private multidimensional artifacts in a forensically-sound manner. Ultimately, these efforts will be applied to enhance LE and private sector investigations so...
as to more efficiently serve and protect society in the face of information age threats.

Erin E. Kenneally

University of California San Diego, Pacific Institute for Computer Security, San Diego Supercomputer Center, 9500 Gilman Dr., La Jolla, CA 92037-0505, USA
Tel.: +1 8585345000.
E-mail address: erin@sdsc.edu

References


5 Erin Kenneally, M.F.S., J.D. is a licensed Attorney who holds Juris Doctorate and Master of Forensic Sciences degrees. Ms. Kenneally consults, researches, publishes, and speaks on prevailing and forthcoming issues at the crossroads of information technology and the law. This includes evidentiary, procedural, and policy implications related to digital forensics, information security and privacy technology. She has lectured and helped coordinate training conferences for officers of the court, law enforcement, and industry professionals concerned with digital evidence and information forensics. She is a Cyber Forensics Analyst at the San Diego Supercomputer Center, liaises and holds leadership positions with the Computer and Technology Computer High Tech Task Force (CATCH) and the Global Privacy and Information Quality Working Group, and provides thought leadership to numerous private and government advisory committees engaged in information technology law issues.